US EPA ARCHIVE DOCUMENT

US EPA ARCHIVE DOCUMENT

Chapter 8
Statistical Analyses

8.1 Overview of Data Analysis

Data for a variety of parameters were available for statistical analysis. These data included the following:

- Concentrations of target pollutants in environmental samples collected at homes and day care centers. Environmental samples included indoor and outdoor air (ng/m^3), soil (ng/g), indoor floor dust collected via HVS3 vacuum (ng/m^2 and ng/g), and drinking water (ng/mL; atrazine only). For homes with recent pesticide applications, concentration data were available for dust collected via wipes from hard floors and food preparation surfaces (ng/m^2) and for transferable residues collected from floors via PUF roller (ng/m^2). Concentration data in dust collected via wipes from hard floors were also available for some locations that did not have carpeted floors from which dust could be collected via HVS3 vacuum.
- Concentrations of target pollutants in personal samples collected from children and adults. Personal samples included duplicate diet solid food samples (ng/g), duplicate diet liquid food samples (ng/mL), and hand wipes (ng/m^2). Adult food samples were analyzed only for selected acid pollutants.
- Information on characteristics, time spent at various locations, and activity patterns associated with the participating children and adults during the sampling period.
- Concentrations of selected acid pollutants and metabolites in urine samples collected from the participating children and adults (ng/mL and µmoles/mole creatinine). For both North Carolina (NC) and Ohio (OH), these pollutants and metabolites included 2,4-D, 1-hydroxybenz[*a*]anthracene, 3-hydroxychrysene, pentachlorophenol, and 3,5,6-TCP. For OH, seven additional metabolites were measured: 3-hydroxybenz[*a*]anthracene, 3-hydroxybenzo[*a*]pyrene, 6-hydroxychrysene, 6-hydroxyindeno[1,2,3-*cd*]pyrene, 1-hydroxypyrene, IMP, and 3-phenoxybenzoic acid.

Pollutant concentrations in multimedia samples (e.g., air, dust, soil, food) were combined with information on activity patterns and physiological parameters to estimate daily potential exposure and absorbed dose for each participant by each of three exposure routes: inhalation, dietary ingestion, and indirect ingestion.[1] ***Potential exposure***, expressed in ng/day and pmoles/day, is defined as the total amount of a pollutant that an individual comes in contact with over a 24-h period. Potential exposure is a route-specific parameter that was calculated from the

[1] Potential exposure and absorbed dose were not estimated for the dermal exposure route due to the limited availability of adequate methods and sufficient background data in the literature.

measured concentrations in those exposure media (multimedia samples) that were relevant to the given exposure route, along with the estimated contact rates with those media. ***Potential absorbed dose***, expressed in ng/kg/day and pmoles/kg/day, is defined as the total dose that could be absorbed into the body over a 24-h period, relative to the participant's body weight. For each exposure route, potential absorbed dose was estimated by assuming a 50% absorption rate for all pollutants and participants (17). This was a conservative approach and was adopted due to the lack of sufficient information available in the scientific literature for most CTEPP target pollutants on the nature of their absorption into the body. Future research may allow these results to be updated be performed on these data when more detailed and accurate absorption rate information becomes available for certain pollutants. For a given study participant, pollutant, and exposure route, potential exposure and potential absorbed dose were calculated if the criteria specified in Section 8.4 were achieved. Section 8.4 provides the detailed formulas that were used to calculate potential exposure and potential absorbed dose.

Aggregate potential exposure and ***aggregate potential absorbed dose*** were defined as the sums of the estimated potential exposure and potential absorbed dose, respectively, across all three exposure routes. Aggregate potential exposure and absorbed dose were calculated for the following eight pollutants and metabolites that were frequently detected (at or above 50%) in several types of multimedia: bisphenol-A, chlorpyrifos, diazinon, di-*n*-butylphthalate, 2,4-D, *cis*-permethrin, *trans*-permethrin, and 3,5,6-TCP.

The concentrations of several parent compounds or their metabolites (specified above) were measured in the urine of children and adults over the 48-h sampling period. Urine samples were combined spot samples rather than total void samples. This was done primarily to prevent placing undue burden on the participants if total void samples were to be collected across the 48-h sampling period. While using spot urine samples rather than total void samples has some limitations (e.g., not allowing for total volume over the 48-h period to be known), a steady-state assumption was made which implied that exposures were chronic in nature. This assumption was reasonable given that information on individual half-lives of the pollutants were unknown, pesticide applications were infrequent, and measured exposures tended to be low. The estimated aggregate potential exposures and absorbed doses of the children were compared with the concentrations of these pollutants in their urine.

Monitoring data were available from a probability sample of 129 children and 129 adults in North Carolina (NC) and a probability sample of 127 children and 127 adults in OH. It is important to note that the study design only permits the outcome of the statistical analyses to be used to characterize the subpopulation of children who reside in the selected counties and who participated in the CTEPP study. The results should not be used to make inferences on larger populations of children, such as all children "in NC, OH, or in the United States," "in low-income and middle/high-income families,"or "in day care centers." Neither can the study design permit results to be used to test hypotheses such as whether exposures differ significantly between all NC children and all OH children. For this report, the statistical summaries and analysis did not consider sample weights assigned to the study participants that would have allowed the results to represent larger populations of children. Future analyses could be

US EPA ARCHIVE DOCUMENT

performed which calculate and take into account sampling weights, from which inferences could be drawn for the populations from which the participants were randomly recruited, namely, preschool children and their caregivers in the randomly-selected counties in NC and OH.

Statistical analyses were conducted to meet each of the four goals detailed in Table 8.1.1. Sub-goals are provided for three of the four goals. Table 8.1.1 also provides an overview of the types of statistical analyses used to address each goal or sub-goal. Details on the statistical analysis approaches are given in Section 8.5.

8.2 Preparation for Statistical Analysis

To prepare for the statistical analyses, several preliminary operations were performed on the collected study data:

- Because high and variable concentrations of selected pollutants were observed in some of the blank samples, it was necessary to apply a background correction to the measured concentrations for these pollutants in some matrices. Background correction to measured concentrations were performed in the following instances:

  - for benzylbutylphthalate and di-*n*-butylphthalate in all sample media collected in both states,
  - for bisphenol-A in dust wipe samples collected in NC, and
  - for *cis*- and *trans*-permethrin in air samples collected in OH.

  The following procedure was used to correct for background contamination. For a given pollutant and matrix, a t-test was applied to the blank data to determine if the mean blank value was significantly different from zero. The mean blank value and an upper 95% confidence bound on the mean were calculated. Then, background-corrected results were calculated by subtracting the mean value adjusted for sample volume, amount, or area (whichever is relevant for the given sample media).
- Sample results labeled as “not detected” were replaced by the method detection limit (MDL) divided by the square root of two for all media except liquid food samples. The pollutant concentrations detected in the liquid food samples were generally very low. When pollutants were detected in liquid food samples at levels close to the MDL, the signal-to-noise ratios for the chromatograms were greater than three. Therefore, not-detected results for the liquid food samples were replaced by the MDL divided by ten.

- In the database, the concentrations of pollutants in dermal wipes were given in ng/sample. Prior to statistical analyses, this value was converted to a loading (ng/m^2 of skin wiped). For each study participant, a tracing of one hand was taken on a sheet of paper, and this tracing was cut out and weighed (in grams). The following equation was then used to calculate the dermal wipe loading (ng/m^2):

US EPA ARCHIVE DOCUMENT

Table 8.1.1 Study Goals and the Statistical Analysis Approaches Used to Address Each Goal

<table>
<tr><th>Study Goal (and Sub-goals)</th><th>Overview of Statistical Analysis Approach</th></tr>
<tr><td>Goal 1: To measure the concentrations of pesticides and other persistent and non-persistent organic pollutants in multimedia at the homes and day care centers of a set of preschool children in several North Carolina and Ohio counties:

<u>Sub-goal 1.1</u>: To quantify the distribution of target pollutants in multimedia (environmental and personal) samples collected from homes and day care centers.

<u>Sub-goal 1.2</u>: To determine on average how multimedia concentrations differ between
– urban and rural environments
– low-income and middle/high-income environments
– microenvironments (home for families with stay-at-home children, home for families with day care children, and day care centers).</td><td><u>Sub-goal 1.1</u>: The following descriptive statistics were calculated on the analytical measurements: sample size, mean (arithmetic and geometric), standard deviation (for untransformed and log-transformed data), percentage detected, minimum reported value, maximum reported value, and selected percentiles (25th, 50th, 75th, 95th). Boxplots of the observed data were also prepared. (See Section 8.5.1)

<u>Sub-goal 1.2</u>: Mixed model analysis of variance (ANOVA) was performed on log-transformed analytical measurements, with the model including fixed effects of income status, urbanicity, and environment type and taking into account correlation in measurements for samples taken within the same day care center. F-tests performed on the model's fixed effects were used to make the statistical comparisons of interest. Results were reported as ratios of geometric means along with 95% confidence intervals, and t-tests were performed to determine whether a particular ratio was significantly different from one. (See Section 8.5.2.1)</td></tr>
<tr><td>Goal 2: To quantify the distribution of child characteristics, activities, and locations that are important for exposure.</td><td>Summary statistics (mean, standard deviation, median, minimum, maximum) were calculated on selected factors that were used to estimate potential exposure levels and potential absorbed dose. These factors included physical characteristics of the study participants (e.g., age, gender, body weight, height, hand surface area), the percentage of time that study participants spent indoors or outdoors at various locations, and the daily amount of solid and liquid food collected from study participants. In addition, the percentage of participating children within specified categories denoting how often certain activities occurred on a daily basis were reported, based upon information obtained from the study questionnaires.</td></tr>
</table>

US EPA ARCHIVE DOCUMENT

Table 8.1.1 Study Goals and the Statistical Analysis Approaches Used to Address Each Goal (cont.)

Study Goal (and Sub-goals)	Overview of Statistical Analysis Approach
Goal 3: To estimate the exposures of the preschool children to these pollutants that they may encounter in their everyday environments: Sub-goal 3.1: To quantify the distribution of potential exposure and potential absorbed dose by exposure route. Sub-goal 3.2: To quantify the distribution of potential exposure and potential dose aggregated over all exposure routes. Sub-goal 3.3: To quantify the distribution of urinary biomarkers concentrations as an indicator of absorbed dose. Sub-goal 3.4: To determine on average how these exposure and dose metrics for each route and aggregated over routes differ between – children in urban and rural settings – children in low and middle/high-income families – day care and stay-at-home children – children and adults in the same household overall – children and adults by stratum.	Sub-goals 3.1 through 3.3: Descriptive statistics were calculated on estimates of potential exposure and potential absorbed dose (by exposure route), aggregate potential exposure, aggregate potential absorbed dose, and urinary biomarker concentrations. Statistics included sample size, mean (arithmetic and geometric), standard deviation (for untransformed and log-transformed data), percentage detected, minimum reported value, maximum reported value, and selected percentiles (25^{th}, 50^{th}, 75^{th}, 95^{th}). Boxplots of the exposure and dose estimates and of the urinary biomarker concentrations were also prepared. *(See Section 8.5.1)* Sub-goal 3.4: Mixed model ANOVA was performed on log-transformed estimates of each of these exposure and dose metrics, as well as on differences in log-transformed estimates between children and adults in the same household. This model included fixed effects of income status, urbanicity, and day care status and took into account correlation in measurements for children attending the same day care center. F-tests performed on the model's fixed effects were used to make the statistical comparisons of interest. Results were reported as ratios of geometric means along with 95% confidence intervals, and t-tests were performed to determine whether a particular ratio was significantly different from one. *(See Section 8.5.2.2)*

US EPA ARCHIVE DOCUMENT

US EPA ARCHIVE DOCUMENT

Table 8.1.1 Study Goals and the Statistical Analysis Approaches Used to Address Each Goal (cont.)

Study Goal (and Sub-goals)	Overview of Statistical Analysis Approach
Goal 4: To apportion the exposures through the inhalation, dietary ingestion, and indirect ingestion routes: Sub-goal 4.1: To estimate the proportion of aggregated potential exposure and absorbed dose that is associated with a given exposure route for the study children, overall and by stratum. Sub-goal 4.2: For each exposure route, determine if this proportion differs for children – in urban and rural settings – from low and middle/high-income families – who attend day care or stay at home. Sub-goal 4.3: Determine whether significant differences exist between exposure routes. Sub-goal 4.4: Characterize how these estimates differ overall between pairs of exposure routes. Sub-goal 4.5: Identify which pairs of exposure routes differ significantly in these estimates.	Sub-goal 4.1: Proportions of aggregated potential exposure and absorbed dose were calculated for each exposure route and analyzed using a logistic regression model that contained effects for income status, urbanicity, and day care status and that accounted for correlation between children attending the same day care center. *(See Section 8.5.2.3, analysis #1.)* Sub-goal 4.2: Wald chi-square tests were performed within the logistic regression to test for significance of the effects in the regression model for a given exposure route to determine whether the proportions differ significantly between two specified groups of children. Estimates of the average proportion within each group and corresponding 95% confidence intervals were reported. *(See Section 8.5.2.3, analysis #1.)* Sub-goal 4.3: Each study participant was represented by a three-dimensional vector of log-transformed potential exposure estimates for the inhalation, dietary, and indirect routes, and a multivariate mixed-model ANOVA was performed on these vectors. This model included fixed effects of income status, urbanicity, and day care status and took into account correlation in measurements for children attending the same day care center, as well as correlation between a participant's three exposure routes. A statistical test performed within this model fit determined whether significant differences existed in the log-transformed exposure or dose estimates among the three routes. *(See Section 8.5.2.3, analysis #2.)* Sub-goals 4.4 and 4.5: Within the multivariate mixed-model ANOVA, pairwise comparisons among the three exposure routes were performed, and these results were reported. *(See Section 8.5.2.3, analysis #2.)*

$$Loading = \frac{A \times D}{4 \times W} \quad (8\text{-}1)$$

where A corresponds to the analytical measurement (ng), D equals the density of the paper on which the hand tracing was made ($\approx$ 80 g/m^2), and W corresponds to the weight of the hand tracing (g). Since the hand wipe involved wiping the front and back of both hands, the reported weight of the hand tracing (W) was multiplied by four within this equation. Note that if a study participant had multiple wipe samples taken at home and/or day care over the 48-h period, the value of A for that participant at a particular location corresponded to the geometric mean of the multiple measurements taken at that location. If W was not reported for a given participant (one in NC, four in OH), then the average value for W was calculated from other participants within the same state and sex category and that were similar in age to the participant, and this average was used to calculate the participant's wipe loading.

- Occasionally, such as when homes did not have carpeted floors or when homes had recent pesticide applications, multiple hard floor surface wipes were collected in the same home. For each of these homes, the geometric mean of these multiple wipe sample results was calculated (after replacing "not detected" values as mentioned above) and used in the statistical analysis. The geometric mean was labeled as "not detected" only when all results used in its calculation were labeled as "not detected."

- A study participant may have had multiple urine samples taken due to recent pesticide application, or a child attending day care may have had urine samples taken both at home and at day care. In these situations, the geometric mean of a participant's urine sample results was calculated and used in the analyses. This geometric mean was labeled as "not detected" only when all results used in its calculation were labeled as "not detected."

- Urine sample concentrations (in both ng/mL and pmoles/mL) were adjusted in two ways: 1) by dividing by the sample's specific gravity, and 2) by dividing by the sample's creatinine level. Creatinine-adjusted urine concentration was expressed in both ng/mg creatinine and µmoles/mole creatinine. Descriptive statistics and statistical analyses were performed on unadjusted and adjusted urine concentrations, for both types of adjustments.

Data labeled as "unusable" by the study's quality control process were not used in statistical summaries and analyses. Measured concentrations were not adjusted based on the recoveries of QC samples (e.g., surrogate recovery samples) prior to including them in summaries or analyses.

US EPA ARCHIVE DOCUMENT

8.3 Strata Considered in the Statistical Analysis

The study goals required the statistical analysis to make comparisons between different strata that were determined according to urbanicity, the income status of the participating families or day care centers, and the type of environment where samples were collected. The different types of statistical analyses required that multimedia sample locations and study participants be stratified. The strata that were considered in the statistical analyses, along with the criteria for placing sampling locations and study participants into strata, were as follows.

- Urban and rural strata: Sampling locations and study participants were placed in the "urban" or "rural" stratum based on the county in which they were located or resided:
  - NC locations and participants were placed in the "urban" stratum if they originated from Buncombe, Durham, Edgecombe, or Mecklenburg counties.
  - OH locations and participants were placed in the "urban" stratum if they originated from Cuyahoga, Franklin, Hamilton, or Licking counties.
  - NC locations and participants were placed in the "rural" stratum if they originated from Jones or Lee counties.
  - OH locations and participants were placed in the "rural" stratum if they originated from Defiance or Fayette counties.

  A county was classified as urban if it contained part of, or was contained within, a Metropolitan Statistical Area (MSA) as defined by the Office of Management and Budget (OMB Bulletin No. 99-04). Counties not meeting this criterion were classified as rural.

- Low-income and middle/high-income strata. Sampling locations from day care centers were placed in the "low income" stratum if the day care center was a Head Start center and in the "middle/high-income" stratum otherwise. Sampling locations from households, as well as all study participants whether stay-at-home or at-day care, were placed in the "low income" stratum if the household's income status (verified during recruitment) achieved the Women, Infants, and Children (WIC) program income guidelines for the period of 7/1/2000 to 6/30/2001, which was equivalent to falling below 185% of the U.S. Poverty Income Guidelines, and were placed in the "middle/high-income" stratum otherwise.

- Children enrolled in day care and children not enrolled in day care. Children were considered enrolled in day care if they attended one of the selected day care centers and were selected to participate based upon meeting all study criteria. Children verified as not attending a day care center or otherwise meeting the day care criteria were labeled as not enrolled in day care.

- Children and adults in the same household. When a child was recruited into the study, a primary caregiver residing in the same household was also identified to participate in the study by providing personal samples (e.g., food, dermal wipes, urine) and activity pattern information needed to calculate potential exposure and potential absorbed dose.

US EPA ARCHIVE DOCUMENT

Table 8.3.1 shows the number of participants in each stratum, for both the NC and OH portions of the study. Because one adult caregiver participated with each child in the study, the number of children and adults in the study was the same within each stratum. While the number of day care and stay-at-home children in the study was similar within each state, the number of participants from urban settings was considerably higher than the number from rural settings. In addition, more middle/high-income households participated in the study compared to low-income households in each state, with the difference in number more apparent in OH. However, a few households in each state did not have sufficient information to allow for their income level to be categorized. Data associated with these households were not included in summaries and statistical analyses when the income status associated with each data value needed to be specified.

Table 8.3.1 Number of Study Participants in Each Stratum, by State

Stratum	Number of Participants			
	North Carolina		Ohio	
	Children	Adults	Children	Adults
Stay-at-Home Child	66	66	69	69
Child Attends Day Care	63	63	58	58
Low-income	59	59	41	41
Middle/High-income	66	66	73	73
Unknown income	4	4	13	13
Urban	108	108	110	110
Rural	21	21	17	17

8.4 Procedures for Calculating Potential Exposure and Potential Absorbed Dose

Estimates of potential exposure were calculated for each study participant under the inhalation, dietary ingestion, and indirect ingestion exposure routes using the equations given below. Estimates of potential exposure via the dermal route were not calculated and were assumed to be negligible. For each participant and exposure route, the potential absorbed dose estimate was calculated as 50% of the potential exposure estimate divided by the participant's body weight (Ross et al., 2001)[1]. Aggregate potential exposure and aggregate potential absorbed dose were defined as the sums of the potential exposure and potential absorbed dose estimates, respectively, across all three exposure routes.

[1] If a participant's body weight was not reported, then the average body weight for other participants within the same state and sex category that were similar in age to the participant was calculated and used in calculating the participant's potential absorbed dose. This approach was necessary for one NC child participant.

US EPA ARCHIVE DOCUMENT

The concentrations of measured pollutants and metabolites in urine over the 48-h sampling period were used as biomarkers of exposure in the study participants. The urinary concentrations of pollutants and metabolites were compared between strata for children and adults.

For each state, Table 8.4.1 lists those pollutants and metabolites that were among those detected in at least 50% of the samples in at least one media type (as seen in Section 9.2) and which were considered for estimating potential exposure and potential absorbed dose in the study participants. Twenty-seven pollutants are listed for NC and 26 for OH. Eight of these pollutants are denoted with an asterisk, as their detection rates were high in multiple media, and some have been commonly found in household consumer products. For these eight pollutants, potential exposure and absorbed dose were estimated in NC and OH children and adults for each exposure route, and aggregate potential exposure and aggregate potential absorbed dose were calculated in these study participants across routes. For the remaining pollutants listed in Table 8.4.1, potential exposure and potential absorbed dose were estimated in children and adults for a given exposure route and state only when the following criteria were satisfied for that pollutant:

- Inhalation route: When at least 45% of the state's indoor air samples, or at least 45% of the state's outdoor air samples, have detected results (i.e., at or above the MDL)
- Dietary ingestion route: When at least 45% of the state's solid food samples, or at least 45% of the state's liquid food samples, have detected results
- Indirect ingestion route: When at least 45% of the state's (vacuum) floor dust samples, or at least 45% of the state's soil samples, have detected results.

Unless otherwise specified, when any of the data entering into the equations below were either not available, could not be assumed to be zero, or were labeled as invalid for a particular study participant, then the potential exposure and potential absorbed dose was not estimated for that participant under the given exposure route, and as a result, aggregate potential exposure and aggregate potential absorbed dose could not be calculated. For purposes of the statistical summaries and analyses, potential exposure level and potential absorbed dose estimates were labeled as "detected" when at least one of the concentrations entering into their calculation was labeled as "detected."

8.4.1 Potential Exposure via Inhalation

Potential exposure via inhalation (ng/day) is a weighted average of measured air concentrations in the different environments in which the participant was present, with the weights corresponding to the time spent in each environment, after adjusting for the participant's estimated ventilation rate:

US EPA ARCHIVE DOCUMENT

US EPA ARCHIVE DOCUMENT

Table 8.4.1 Pollutants Considered for Estimating Potential Exposure and Potential Absorbed Dose for Study Participants in a Given State

Pollutant	NC	OH	Pollutant	NC	OH
Benz[*a*]anthracene	T	T	Dibenz[*a*,*h*]anthracene	T	T
Benzo[*b*]fluoranthene	T	T	Di-*n*-butylphthalate*	T	T
Benzo[*k*]fluoranthene	T	T	*p,p'*-DDE	T	T
Benzo[*ghi*]perylene	T	T	2,4-Dichlorophenoxyacetic acid*	T	T
Benzo[*a*]pyrene	T	T	Heptachlor	T	
Benzo[*e*]pyrene	T	T	Indeno[1,2,3-*cd*]pyrene	T	T
Benzylbutylphthalate	T	T	Pentachlorophenol	T	T
Bisphenol-A*	T	T	*cis*-Permethrin*	T	T
alpha-Chlordane	T	T	*trans*-Permethrin*	T	T
gamma-Chlordane	T	T	PCB 52	T	T
Chlorpyrifos*	T	T	PCB 95	T	T
Chrysene	T	T	PCB 101	T	T
Cyfluthrin	T	T	3,5,6-Trichloro-2-pyridinol*	T	T
Diazinon*	T	T			

* Pollutants for which potential exposure and potential absorbed dose were calculated for each exposure route for the study participants in each state, and for which aggregate potential exposure and aggregate potential absorbed dose were calculated (across exposure routes).

$$Exp_{inh} = \frac{(C_{di} \times t_{di}) + (C_{do} \times t_{do}) + (C_{hi} \times t_{hi}) + (C_{ho} \times t_{ho}) + (C_{away} \times t_{away})}{t_{di} + t_{do} + t_{hi} + t_{ho} + t_{away}} \times V \qquad (8\text{-}2)$$

where the notation is as follows:

C_{di} = Indoor air concentration in the participant's day care center classroom (ng/m^3)
C_{do} = Outdoor air concentration at the participant's day care center (ng/m^3)
C_{hi} = Indoor air concentration in the participant's home (ng/m^3)
C_{ho} = Outdoor air concentration at the participant's home (ng/m^3)
C_{away} = Air concentrations in indoor locations other than the participant's day care center or home where the participant may spend time (ng/m^3)
t_{di} = Time spent indoors at day care when indoor air is being sampled there (hr)
t_{do} = Time spent outdoors at day care when outdoor air is being sampled there (hr)
t_{hi} = Time spent indoors at home when indoor air is being sampled there (hr)

t_{ho} = Time spent outdoors at home when outdoor air is being sampled there (hr)
t_{away} = Time spent indoors at locations other than day care or home during the sampling period (hr)
V = Ventilation rate, estimated as follows from information in the EPA Exposure Factors Handbook:

- 6.8 m^3/day for children less than 36 months of age
- 8.3 m^3/day for children aged 36 months or higher
- 11.3 m^3/day for adult females
- 15.2 m^3/day for adult males

For each of the participating children and their adult caregivers, an air sample was collected over a 48-h period in each of the indoor and outdoor environments at their homes. In addition, an air sample was collected over a 48-h period in each of the indoor and outdoor environments of participating day care centers, with most centers having separate indoor air samples taken in each classroom containing a participating child. Thus, the values of C_{di}, C_{do}, C_{hi}, and C_{ho} for a given participant were taken to be the measured concentrations in the four air samples associated with that participant. However, no air samples were taken in indoor environments other than homes and day care centers to allow C_{away} to be estimated. Thus, to arrive at a value for C_{away}, the median of all indoor air concentration measures taken in a given state was calculated for each pollutant listed in Table 8.4.1, and this median, specified in Appendix F, was taken to be the estimate of C_{away} for each study participant in that state. Equation (8-2) does not include a term for air concentration in outdoor environments away from homes or day care centers, as the times spent in these other outdoor environments were assumed to be trivial (i.e., near zero) for the study participants.

For day care children, values of t_{di} and t_{do} in equation (8-2) were obtained from information recorded on the Child Activity Diary and Food Survey (Form 10), completed by day care teachers. For day care children and their adult caregivers, values of t_{hi} and t_{ho} were obtained from information recorded on Child Activity Diary and Food Survey (Form 9), completed by day care parents), and t_{away} was calculated from information recorded on Forms 09 and 10. For stay-at-home children and their adult caregivers, values of $t_{hi,}$ $t_{ho,}$ and t_{away} were determined from information recorded on Form 08 (Child Activity Diary and Food Survey, completed by "home" parents). For stay-at-home children and all adult caregivers in the study who were not exposed to a day care environment, t_{di} and t_{do} were both set equal to 0.

8.4.2 Potential Exposure via Dietary Ingestion

Potential exposure level via dietary ingestion (ng/day) is a weighted sum of measured concentrations in both solid and liquid food within the day care and home environments in which the participant was present, with each concentration multiplied by the amount of the collected sample (representing the total amount of food eaten by the participant):

US EPA ARCHIVE DOCUMENT

$$Exp_d = [(C_{dl} \times M_{dl}) + (C_{ds} \times M_{ds}) + (C_{hl} \times M_{hl}) + (C_{hs} \times M_{hs})] \times \frac{1}{N_f} \tag{8-3}$$

where the notation is as follows:

C_{dl} = Concentration in liquid food sample collected in the participant's day care classroom (ng/mL)
C_{ds} = Concentration in solid food sample collected in the participant's day care classroom (ng/g)
M_{dl} = Total volume of liquid food sample collected in the participant's day care classroom (mL)
M_{ds} = Total weight of solid food sample collected in the participant's day care classroom (g)
C_{hl} = Concentration in the participant's liquid food sample collected at home (ng/mL)
C_{hs} = Concentration in the participant's solid food sample collected at home (ng/g)
M_{hl} = Total volume of the participant's liquid food sample collected at home (mL)
M_{hs} = Total weight of the participant's solid food sample collected at home (g)
N_f = Number of days over which all food samples (liquid and solid) associated with the participant were collected.

Because each food sample at a given location for a given study participant corresponded to a composite of total food consumed by the participant over a two-day period, the value of N_f was set equal to two for each participant. Participants that drank only water at day care and/or home were assumed to have liquid food sample concentrations (C_{dl} and C_{hl}, respectively) of 0 ng/mL for that environment. Although C_{dl} and C_{ds} were not measured for stay-at-home children and for all adult caregivers, the values of M_{dl} and M_{ds} for these participants were zero, and therefore, these concentrations were not a factor in calculating the potential exposure level.

8.4.3 Potential Exposure via Indirect Ingestion

Potential exposure via indirect ingestion (i.e., ingestion of dust and soil) (ng/day) is a weighted average of measured floor dust and soil concentrations in the indoor and outdoor environments, respectively, in which the study participant was present, with each concentration scaled by the participant's assumed ingestion rate:

$$Exp_n = \frac{(D_{dd} \times M_d \times t_{di}) + (D_{ds} \times M_s \times t_{do}) + (D_{hd} \times M_d \times t_{hi}) + (D_{hs} \times M_s \times t_{ho})}{t_{di} + t_{do} + t_{hi} + t_{ho}} \tag{8-4}$$

where the notation is as follows:

US EPA ARCHIVE DOCUMENT

D_{dd} = Concentration in the day care center/classroom's HVS3 (vacuum) floor dust sample (ng/g)
D_{ds} = Concentration in day care center's play area soil sample (ng/g)
D_{hd} = Concentration in home's HVS3 floor dust sample (ng/g)
D_{hs} = Concentration in home's play area soil sample (ng/g)
M_d = Participant's estimated daily ingestion rate of dust (g/day)
M_s = Participant's estimated daily ingestion rate of soil (g/day)

and t_{di}, t_{do}, t_{hi}, and t_{ho} are defined in the same way as in equation (8-2) (i.e., times spent indoors and outdoors in the day care and home environments). For stay-at-home children and all adult caregivers who were not exposed to a day care environment, t_{di} and t_{do} were both set equal to 0. Any indirect ingestion that might have occurred outside of the day care center and home environments was assumed to be trivial, and therefore, was not included in equation (8-4). Daily ingestion rates of dust and soil were estimated according to the published literature (15-16) and from the collected questionnaire data on children's activity patterns. For participating children, daily ingestion rates were estimated by placing each child into one of three groups (Groups A, B, or C) according to information recorded on study survey forms on how often the child conducted activities that could lead to dust and soil ingestion, such as teething, chewing, and putting objects into his/her mouth. For soil ingestion activity, responses from the following two questions on Form 04 (parent pre-monitoring questionnaire) were evaluated:

(1) Question C5: How often did [the child] play with sand or dirt?
(2) Question C6: Which of the following have you seen your child eat: dirt, sand, snow?

For dust ingestion activity, responses from the following questions on Form 04 were evaluated:

(1) Question C12: Did your child use a pacifier in the past month?
(2) Question C13a: In the past month, did [your child] suck or chew his/her thumb/fingers?
(3) Question C13b: In the past month, did [your child] suck or chew his/her toe/foot?
(4) Question C16: Did [your child] ever put his/her mouth on the floor and lick the floor?
(5) Question C21: Is your child currently teething?
(6) Question C22: How often did [your child] put toys in his/her mouth?
(7) Question C23: Did [your child] put any things other than toys or food in his/her mouth?

Algorithms were established to assign a daily soil ingestion rate and a daily dust ingestion rates to a child based upon the responses to the above questions for that child, with the specific rates that entered into the algorithms being selected in conjunction with the published literature (15-16). Appendix G provides details on these algorithms. Separately for dust and soil ingestion, the algorithms placed children into Groups A, B, or C based upon whether their

US EPA ARCHIVE DOCUMENT

activity levels were considered high, medium, or low, respectively. For both dust and soil, daily ingestion rates were assigned as follows:

- Children in Group A: Daily ingestion rate = 0.100 g/day
- Children in Group B: Daily ingestion rate = 0.050 g/day
- Children in Group C: Daily ingestion rate = 0.025 g/day

For all participating adult caregivers, assigned ingestion rates were M_d=25 mg/day for dust and M_s=50 mg/day for soil. Note that while the activity diaries and questionnaires provide useful information for exposure assessment, they were not fully validated prior to their use in this study.

8.5 Statistical Analysis

This section details the methods associated with the statistical summaries and analyses that were applied to the study data in order to address each of the study's goals and sub-goals. The data were prepared for analysis as discussed in Section 8.2, then were statistically summarized and analyzed using Version 8 (Release 8.2) of the SAS® System. These statistical methods were applied independently to data from NC and OH.

8.5.1 Descriptive Statistics

As mentioned in Table 8.1.1, descriptive statistics were generated on the study data in order to address the following five goals or sub-goals:

- Sub-goal 1.1: to quantify the distribution of target pollutants in multimedia samples at homes and day care centers
- Goal 2: to quantify the distribution of child characteristics, activities, and locations that are important for exposure
- Sub-goal 3.1: to quantify the distribution of potential exposure and potential absorbed dose by exposure route
- Sub-goal 3.2: to quantify the distribution of aggregate potential exposure and potential absorbed dose
- Sub-goal 3.3: to quantify the distribution of urinary biomarker concentrations as an indicator of absorbed dose.

The SAS® System's UNIVARIATE procedure was applied to the relevant study data to calculate the descriptive statistics. For Goal 2, the list of summarized parameters and the descriptive statistics calculated on these parameters were given in Table 8.1.1. For the four sub-goals, the descriptive statistics included the sample size, mean (arithmetic and geometric), standard deviation (for untransformed and log-transformed data), percent of results labeled as detected, minimum reported value, maximum reported value, and selected percentiles of the observed data distribution (25^{th}, 50^{th}, 75^{th}, 95^{th}). Means and standard deviations were reported only when at least 50% of the data entering into their calculation were detected. A given percentile was

US EPA ARCHIVE DOCUMENT

reported only when the observed data values at the percentile exceeded the MDL. The maximum reported value was reported only when at least one detected measurement was reported, and the minimum reported value was reported only when 100% of the reported measurements were detected. These descriptive statistics are included as appendices to this report.

Also, for the four sub-goals specified above, boxplots were prepared which portrayed the distribution of observed data values as a box-type diagram, within which the 25th, 50th, and 75th percentiles, the geometric mean, and the range of the data were expressed graphically. Details on how to interpret the boxplots are given in Section 9.3.1.

8.5.2 Analysis of Variance (ANOVA) Modeling

Model-based analysis of variance (ANOVA) methods were applied to the study data in order to address Sub-goal 1.2, Sub-goal 3.4, and Goal 4, as detailed in the three subsections below. In each case, the ANOVAs were repeatedly applied to different subsets of study data using the SAS® System's MIXED and GENMOD procedures, with each subset of data associated with a specific target pollutant and media type/dose metric. While the ANOVA approach applies when the data used in the analysis satisfies certain statistical assumptions, the same approach was applied to each subset of data (i.e., each combination of pollutant and sample type) when addressing a particular study goal. This was done in order to maintain consistency in approach across the repeated analyses, so that the outcomes of the analyses could be more comparable across the pollutants and sample types. Note that the outcome of statistical analyses of urine, potential exposure, and potential absorbed dose data was not affected by whether the data were expressed in mass concentration or molar concentration units.

8.5.2.1 Sub-goal 1.2: To determine on average how multimedia concentrations differ between urban and rural environments, low-income and middle high-income environments, and microenvironments

Multimedia (environmental and food) samples were collected at the homes and day care centers of the participating children. Within a day care center, indoor environmental samples were linked to children by classroom. These two locations, along with an indicator of whether or not a child attended day care, defined three possible *microenvironments*: 1) the day care microenvironment; 2) the home microenvironment for stay-at-home children, and 3) the home microenvironment for children attending day care. Additionally, multimedia samples were classified by income status (low or middle/high) and urbanicity (urban or rural) according to the microenvironment from which they were collected. The primary aim of the data analysis was to make statistical comparisons among the three microenvironments, although comparisons were also made according to income status and urbanicity.

For a given multimedia sample type and pollutant (with the exception of dermal wipes), let Y_{ijk} denote the log-transformed analytical measurement associated with a sample collected in the ith environment type, where the sample is identified as follows:

US EPA ARCHIVE DOCUMENT

US EPA ARCHIVE DOCUMENT

- For samples collected in a day care center environment (i=1), the sample taken in the jth classroom within the kth day care center in the study.
- For samples collected in the home environment of a stay-at-home child (i=2), the sample collected in the kth home of this type in the study. (Here, j is assumed to be equal to one as only one sample was taken per home).
- For samples collected in the home environment of a day care child (i=3), the sample taken in the kth home of this type in the study. (Here, j is assumed to be equal to one as only one sample was taken per home.)

Then, for a particular combination of pollutant and environmental/food sample type, the following analysis of variance (ANOVA) model was applied to the log-transformed analytical measurements Y_{ijk}:

$$Y_{ijk} = \mu + \eta_i + \gamma_1 M_{ik} + \gamma_2 U_{ik} + \delta_k + \varepsilon_{ijk} \tag{8-5}$$

where

μ = an overall constant,
η_i = effect of originating from the ith environment type,
γ_1 = effect of originating from a middle/high-income environment versus a low income environment
M_{ik} = indicator of income status associated with the kth day care center or home within the ith environment type (i.e., M_{ik}=1 if middle/high-income and =0 if low income),
γ_2 = effect of originating from an urban environment versus a rural environment
U_{ik} = indicator of urbanicity associated with the kth day care center or home within the ith environment type (i.e., U_{ij}=1 if an urban area and =0 if a rural area),
δ_{jk} = a random term corresponding to the kth home or day care center, and
ε_{ijk} = a random error term representing random variation not explained by the model.

Because no interactions are included in the model, any interaction effects are included in the random error term. The variance-covariance matrix of δ_k was defined to account for correlation in measurements for samples taken in different classrooms (j) within the same day care center (k), while the variance-covariance matrix of ε_{ijk} was defined under the assumption that the values of ε_{ijk} for different samples are independent.

The statistical significance of environment type (η_i), income status (γ_1), and urbanicity (γ_2) on the value of Y_{ijk} was determined by applying F-tests within the ANOVA, and significance levels of these F-tests were reported. When the F-test for the effect of environment type (η_i) was found to be significant at the 0.05 level and all three environment types were represented by the data, multiple comparisons (using Tukey's studentized range test) were performed to identify which of the three pairs of environment types differed significantly, and the significance levels (adjusted for the multiple comparisons) associated with each of the three pairs were reported. Additionally, a t-test was performed within the ANOVA to determine if the day care environment differed significantly with the mean of the two home environment types, and the significance level of this test was also reported.

To characterize how the analytical measurements differ between two strata (e.g., urban vs. rural, low income vs. middle/high-income), the ANOVA model was used to estimate the average log-transformed analytical measurement ("least squares mean") for each stratum. Then, the difference in the least squares means of the two strata was calculated, a t-test was performed within the ANOVA to determine whether this difference was statistically significant at the 0.05 or 0.01 levels, and a 95% two-sided confidence interval on this difference was also calculated within the ANOVA. The estimated difference in least squares means and its 95% confidence interval were then exponentiated, resulting in a ratio of estimated geometric means between the two strata and a corresponding 95% two-sided confidence interval on this ratio. The estimated ratio, its 95% confidence interval, and the outcome of the statistical test for significant difference between the two strata were reported.

Because a statistical comparison between home and day care environments was also of interest, a linear contrast was constructed within the ANOVA to estimate the difference in average log-transformed measurements between these two environments. Because the home environment consisted of two of the three microenvironments (i.e., the home environment for day care children and the home environment for stay-at-home children), the linear contrast was specified as the average log-transformed analytical measurement for the day care microenvironment, minus the average of the average log-transformed analytical measurements associated with the two home microenvironments. As with the other comparisons of strata, a t-test was performed within the ANOVA to determine whether this difference between home and day care environments was significant at the 0.05 or 0.01 levels, and a 95% two-sided confidence interval on this difference was calculated within the ANOVA. A ratio of estimated geometric means between the home and daycare environments was also calculated, along with a 95% two-sided confidence interval on this ratio.

While all pollutants were considered in the analysis of environmental sample data, model (8-5) was applied to only those combinations of pollutant and multimedia samples that met the following two criteria:

- At least 50% of the values of Y_{ijk} were labeled as detected.
- Values of Y_{ijk} were available for at least two of the three environment types.

Within an application of the analysis, if data were available from only one of a given microenvironment (e.g., data were available for only one day care center), then data for that microenvironment were excluded from that application of the analysis. The check for whether at least 50% of the values were detected occurred after any necessary data exclusions were made.

For the adult food sample type, microenvironments were relevant based upon whether or not their child attended day care: home microenvironment for stay-at-home children (i=2), and home

US EPA ARCHIVE DOCUMENT

microenvironment for day care children (i=3). This is because all adult-specific data were collected within the home microenvironment.

A slightly different ANOVA model was used for analysis of dermal wipe data. Dermal wipes were collected for each study participant (child and adult) at their home and, for day care children, at their day care center. Thus, day care children could have up to two dermal wipe measurements, corresponding to their home and day care microenvironments. The statistical analysis of dermal wipe data, therefore, needed to take into account correlation in the day care and home dermal wipe samples for day care children. In the analysis of dermal wipe data, let Y_{ijk} denote the log-transformed analytical measurement associated with a dermal wipe sample collected in the i^{th} environment type, where the sample is identified as follows:

- For day care children, the sample taken in the i^{th} environment (day care [i=1] or home [i=3]) from the j^{th} child enrolled in the k^{th} day care center of the study.
- For stay-at-home children and for all adult participants, the sample collected in the k^{th} home of the environment type determined by whether or not the child attends day care (i=2 or 3). (Here, j is assumed to be equal to one as only one child and one adult participated from each home.)

The ANOVA model applied to the dermal wipe sample data took the following form:

$$Y_{ijk} = \mu + \eta_i + \gamma_1 M_{ij} + \gamma_2 U_{ij} + \delta_k + \varepsilon_{ijk} \qquad (8\text{-}6)$$

where the terms are as defined for equation (8-5) except for the following:

M_{ij} = indicator of income status associated with the j^{th} study participant within the i^{th} environment type (i.e., M_{ij}=1 if middle/high-income and =0 if low income),
U_{ij} = indicator of urbanicity associated with the j^{th} study participant within the i^{th} environment type (i.e., U_{ij}=1 if an urban area and =0 if a rural area),

Because no interactions are included in the model, any interaction effects are included in the random error term (ε_{ijk}). The variance-covariance matrix of δ_k was defined to account for correlation in measurements for samples taken from different children (j) within the same day care center (k), while the variance-covariance matrix of ε_{ijk} was defined to account for correlation in measurements for samples taken from the same child (j) at different environment types (i) (i.e, day care and home).

The results for the tests of significance for environment, urbanicity, and income status on the log-transformed analytical measurement, and their estimated geometric ratios and associated 95% confidence intervals, were reported in the same manner as for the environmental/food samples. Model (8-6) was fitted separately for each pollutant, as well as separately for adults and children.

US EPA ARCHIVE DOCUMENT

8.5.2.2. Sub-goal 3.4: To determine on average how potential exposure and absorbed dose metrics for each route and aggregated over routes differs between children in urban and rural settings, children in low and middle/high-income settings, day care children and stay-at-home children, and children and adults by stratum

The analysis approach presented in this subsection was performed on the potential exposure and absorbed dose estimates for the target pollutants listed in Table 8.4.1, when the data for these pollutants achieved the criteria specified in Section 8.4 for a given exposure route. The analyses were executed separately for each exposure route. In addition, this approach was performed on urine concentration data (both adjusted and unadjusted for specific gravity and creatinine concentration), separately for each pollutant measured in urine, and on aggregated potential exposure level and aggregated potential absorbed dose estimates, separately for each of the eight pollutants labeled with asterisks in Table 8.4.1.

Let j denote a specific household enrolled in the study. The analyses addressing Sub-goal 3.4 were performed on the measures Y_j, with separate analyses being conducted by pollutant and for each of the following definitions of Y_j:

- Log-transformed potential exposure level for the child in the j^{th} household (separate analyses by exposure route)
- Log-transformed potential absorbed dose for the child in the j^{th} household (separate analyses by exposure route)
- Log-transformed aggregated potential exposure level for the child in the j^{th} household
- Log-transformed aggregated potential absorbed dose for the child in the j^{th} household
- Log-transformed unadjusted urine concentration for the child in the j^{th} household
- Log-transformed urine concentration, adjusted for specific gravity, for the child in the j^{th} household
- Log-transformed urine concentration, adjusted for creatinine, for the child in the j^{th} household
- Difference in log-transformed potential exposure level between the child and adult in the j^{th} household (separate analyses by exposure route)
- Difference in log-transformed potential absorbed dose between the child and adult in the j^{th} household (separate analyses by exposure route)
- Difference in log-transformed aggregated potential exposure level between the child and adult in the j^{th} household
- Difference in log-transformed aggregated potential absorbed dose between the child and adult in the j^{th} household
- Difference in log-transformed unadjusted urine concentration between the child and adult in the j^{th} household
- Difference in log-transformed urine concentration, adjusted for specific gravity, between the child and adult in the j^{th} household

US EPA ARCHIVE DOCUMENT

- Difference in log-transformed urine concentration, adjusted for creatinine, between the child and adult in the j^{th} household.

The ANOVA model applied to data for a given combination of pollutant and Y_j definition was the following:

$$Y_j = \mu + \gamma_1 M_j + \gamma_2 U_j + \gamma_3 D_j + \varepsilon_j \tag{8-7}$$

where

μ = an overall constant,
γ_1 = effect of a middle/high-income household versus a low income household,
M_j = indicator of the j^{th} household's income status (M_j=1 if middle/high-income, =0 if low income),
γ_2 = effect of an urban household versus a rural household,
U_j = indicator of the j^{th} household's urbanicity (U_j=1 if urban, =0 if rural),
γ_3 = effect of a child enrolled in day care versus staying at home,
D_j = indicator of child's day care status in the j^{th} household (D_j=1 if day care, =0 if non-day care), and
ε_j = a random error term representing random variation not explained by the model.

The variance-covariance matrix of ε_j was defined to account for correlation in measurements among households whose children attend the same day care center.

In a given fitting of model (8-7), the statistical significance of urbanicity, income status, and day care status on the value of Y_j was determined by testing for the significance of their corresponding coefficients in the model using F-tests and reporting the significance levels of these tests. As in the previous models, because no interactions of these factors are included in the model, only the main effects of these factors were tested. Thus, any interaction effects are included in the model's random error term.

When the definition of Y_j corresponded to some child-specific measure (i.e., not a child vs. adult difference), the ratio of estimated geometric means between two strata (e.g., urban vs. rural, low income vs. middle/high-income, day care vs. non-day care) were reported for this measure as in the previous models, along with 95% two-sided confidence intervals. T-tests were also performed to determine whether a particular ratio was significantly different from one, implying no significant difference between the two strata represented by the ratio. When the definition of Y_j corresponded to a difference in measures between children and adults within the same household, the ratio of estimated geometric means for children versus adults in the same household were reported overall and for each stratum, along with 95% two-sided confidence intervals. In addition, a one-sided t-test was performed within the model fitting that tested whether, overall, children tended to have significantly higher measures than their adult caregivers. For the individual strata, two-sided t-tests were performed to test whether children's measures differed significantly from their adult caregivers.

US EPA ARCHIVE DOCUMENT

8.5.2.3 Goal 4: To apportion the exposures through the inhalation, dietary ingestion, and indirect ingestion routes

For the eight pollutants highlighted in Table 8.4.1 for which aggregated potential exposure level and aggregated potential absorbed dose were estimated, this goal focuses on characterizing how these aggregated estimates were apportioned across the three exposure routes considered in this study (inhalation, dietary ingestion, and indirect ingestion) and noting which routes were more important contributors to aggregate potential exposure or aggregate potential absorbed dose than others. As indicated in Table 8.1.1, this goal was divided into the following five sub-goals:

- 4.1 To estimate the proportion of aggregated exposure and dose that is associated with a given exposure route for the study children overall and by stratum.
- 4.2 For each exposure route, determine if this proportion differs for children
  - a. in urban and rural settings
  - b. from low and middle/high-income families
  - c. who attend day care or stay at home
- 4.3 Determine whether significant differences exist between exposure routes
- 4.4 Characterize how these estimates differ overall between pairs of exposure routes
- 4.5 Identify which pairs of exposure routes differ significantly in these estimates

To address each of these sub-goals, two types of analyses were developed and executed:

- Analysis #1 (Sub-goals 4.1 and 4.2): Characterizes the proportion of the aggregated value that is associated with a specific exposure route, both overall and by stratum, and determines whether these proportions differ significantly between strata. This analysis was performed separately by pollutant and exposure route.

- Analysis #2 (Sub-goals 4.3, 4.4, and 4.5): Compares average log-transformed measures between exposure routes. This analysis was performed separately by pollutant and for potential exposure and potential absorbed dose.

Each of these analysis approaches is now discussed.

Analysis #1. When applied to a given exposure route, this analysis involved calculating p_j, or the proportion of the estimated aggregated exposure that is associated with the given exposure route, for the j^{th} participant. To make statistical comparisons of the value of p_j between strata, the following logistic regression model was used:

$$\log(p_j/(1-p_j)) = \mu + \gamma_1 M_j + \gamma_2 U_j + \gamma_3 D_j + \varepsilon_j \tag{8-8}$$

US EPA ARCHIVE DOCUMENT

where the terms in this model are as defined for equation (8-7). Generalized estimating equations were used to allow values of the proportion p_j associated with children enrolled in the same day care center to be correlated.

The presence of significant differences among strata was determined by testing the statistical significance of the corresponding model coefficients via a Wald chi-square test. For example, the differences of the proportion between children living in urban areas and children living in rural areas was investigated by testing for the significance of the γ_2 coefficient in model (8-8). Significance levels of tests for significant differences between urban and rural strata, between middle/high and low income strata, and between day care and non-day care strata were reported, along with estimates and corresponding 95% confidence intervals for the average proportion for each stratum. The estimated average proportion for each stratum was determined by solving model (8-8) for the value of p_j for the given stratum (i.e., calculating the inverse logit).

Because the proportion p_j is calculated for each participant for a given exposure route, the outcome of this calculation is the same whether potential exposure level or potential absorbed dose is used. This is because the absorption rate (50%) and the participant's body weight cancel out from the numerator and denominator of the proportion equation. Thus, for a given exposure route, only one analysis was necessary between these two endpoints.

Analysis #2. To investigate whether potential exposure level or potential absorbed dose differed significantly among the three exposure routes and among strata, this analysis involved a multivariate ANOVA fitted to the log-transformed estimates for a given pollutant. This approach is similar to that discussed in Section 8.5.2.2, except the model is multivariate in nature in that it is applied to the vector of three log-transformed estimates associated with each exposure route. For the i^{th} entry (or exposure route) in this vector (i=1, 2, 3), the multivariate ANOVA model is as follows:

$$Y_{ij} = \mu + \gamma_1 M_j + \gamma_2 U_j + \gamma_3 D_j + \delta_j + \varepsilon_{ij} \qquad (8\text{-}9)$$

where

Y_{ij} = log-transformed exposure or dose estimate for the j^{th} study participant via the i^{th} exposure route,

μ = an overall constant,

γ_1 = effect of a middle/high-income household versus a low income household,

M_j = indicator of the household income status for the j^{th} study participant (M_j=1 if middle/high-income, =0 if low income),

γ_2 = effect of an urban household versus a rural household,

U_j = indicator of the urbanicity of the household containing the j^{th} study participant (U_j=1 if urban, =0 if rural),

γ_3 = effect of a child enrolled in day care versus staying at home,

US EPA ARCHIVE DOCUMENT

D_j = indicator of child's day care status in the household containing the j^{th} study participant (D_j=1 if day care, =0 if non-day care),

δ_j = random day care center effect, which accounts for correlation between children attending the same day care center, and

ε_j = a random error term representing random variation not explained by the model that accounts for correlation between exposure routes for each participant.

When fitting model (8-9), a statistical test was performed to determine whether significant differences existed in the log-transformed exposure or dose estimates among the three exposure routes. Then, pairwise comparisons among the three exposure routes were performed, and the results were reported. In addition, the estimated ratio of geometric means between two exposure routes were calculated and reported for each pair of routes, along with a 95% confidence interval on the ratio.

US EPA ARCHIVE DOCUMENT